

NexGen FOIA Technology Showcase

Day 1, Topic 3: Redaction tools (including tools that allow for automatic redaction of similar forms and record types including paper, digital content, videos, data, etc.)

Since its inception in 1967, the Freedom of Information Act/Privacy Act (FOIA/PA) has placed the onus on government agencies to disclose information and release documents upon request by the public. Agencies are facing a growing demand from its citizens to respond to these requests with speed and accuracy. In its current state, responding to FOIA/PA requests require large amounts of manual and time-intensive labor, cumulating in an increasing backlog of unprocessed requests. Fulfilling these requests is becoming increasingly expensive and directly contributes to ballooning overhead. Agencies require automated processing support services to close the growing capacity gap of documents that require manual processing for case management, redaction and eDiscovery.

IBM brings extensive experience deploying solutions, at scale, within Government environments in conjunction with unique past performance successfully delivering the largest Intelligent Automation Platform (IAP) in production within the Government. IBM is uniquely positioned to support our Federal Partners in closing this gap through the implementation of modern automation processes and technology to more efficiently process NARA's FOIA requests while maintaining a human-centric and quality management focus. IBM has successfully established partnerships to apply technologies that construct our automation platform, such as Robotic Process Automation (RPA), Intelligent Character Recognition (ICR), Intelligent Form Recognition (IFR), Natural Language Processing (NLP), and Artificial Intelligence (AI) to automate document intake, redaction, and eDiscovery processes. In this capacity, IBM has worked side-by-side with Government agencies to start-up and scale programs and automation platforms that operate 24x7.

IBM's automation platform is an Authorized To Operate (ATO) Commercial off-the-shelf (COTS) solution that has been successfully deployed across development, test, and production environments within the Government; is operable across security levels and agnostic to cloud environments; and is integrated with backend databases and incoming document ingestion streams. We bring processes and procedures to facilitate continuous performance monitoring and human-in-the-loop interaction to improve processing quality and accuracy.

Automated Redaction

Over the last 14 months, IBM supported the Veterans Benefits Administration in deploying IBM's IAP to assist them with handling a backlog of Freedom of information Act and Privacy Act (FOIA/PA) requests. IBM and VBA developed a largely unattended automation capability for processing FOIA/PA requests. Our platform leverages COTS for case management, RPA, and AI-enabled OCR and auto-redaction engines to provide streamlined workflows and automated processing of documents. The redaction workflow is depicted in Figure 1 and key capability areas outlined as follows.

Extract – Before we extract information from documents, we need to verify that appropriate documents are **retrieved** from system of record. The team accomplishes this through automatic API/RPA retrieval of suggested response documents based on requester and request type. To handle document retrieval and redaction at large scale, our platform also requires coordinated control and orchestration. During **extraction**, our team uses a combination of OCR/ICR

extraction, as well computer vision and AI to extract relevant information from documents. This is a critical step in the process where we transform unstructured content into usable content for use downstream in the process.

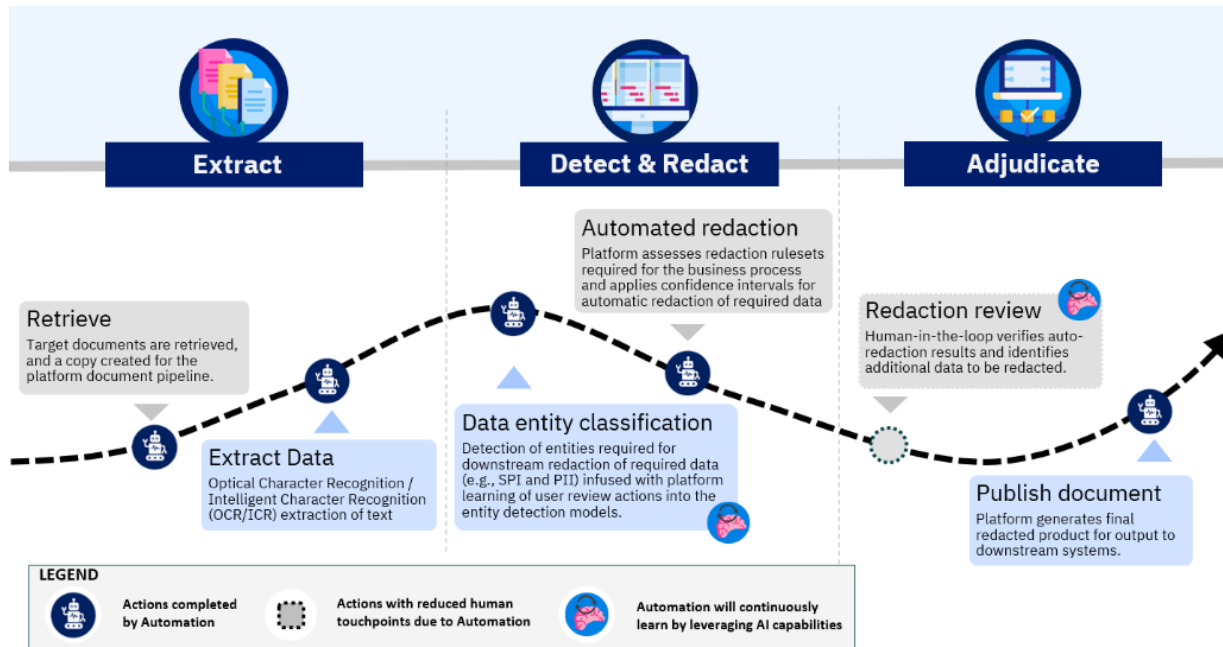


Figure 1: Redaction on IBM Intelligent Automation Platform

Detect and Redact – Once we have successfully extracted information from the documents, it is up to our data entity **detection and classification** to parse through data received from OCR/ICR, and identify appropriate entities that are needed for redaction. Entity detection and classification may look different for each organization, as custom models will need to be built for specific organizational business processes. The final step is to perform the **Redaction** of the correct information. As was the case with detection and classification, custom redaction rulesets will need to be built for the business process. Redaction applies confidence intervals for automatic redaction of required data.

Adjudicate – Based on quality requirements, strategically placed human-in-the-loop (HIL) workflow steps will need to be inserted to validate the redactions made by the auto-redaction process. This **adjudication** step serves as a validation and quality assurance for the customer, as well as a feedback loop for intelligent entity detection.

Submitted by: IBM Corporation, 6710 Rockledge Drive, Bethesda, MD 20817
To Learn More: Bridget Deary, bmwalsh@us.ibm.com,
Patrick McDonnell, pat.mcdonnell@us.ibm.com